A Comprehensive Study of the Effect on Colorectal Cancer Survival of Common Germline Genetic Variation Previously Linked with Cancer Prognosis

Title: A comprehensive study of the effect on colorectal cancer survival of common germline genetic variation previously linked with cancer prognosis.

Authors: Yazhou He[1,2,3], Maria Timofeeva[1,2], Xue Li[3], Farhat VN. Din[1,2], James P Blackmur[2], Peter Vaughan-Shaw[2], Victoria Svinti[2], Susan M. Farrington[1,2], Harry Campbell[3], Malcolm G. Dunlop[1,2*], Evropi Theodoratou[1,3*]

Author affiliations:

1. Cancer Research UK Edinburgh Centre, Medical Research Council Institute of Genetics & Molecular Medicine, Western General Hospital, The University of Edinburgh, Edinburgh, UK
2. Colon Cancer Genetics Group, Medical Research Council Human Genetics Unit, Medical Research Council Institute of Genetics & Molecular Medicine, Western General Hospital, The University of Edinburgh, Edinburgh, UK
3. Centre for Global Health Research, Usher Institute, The University of Edinburgh, Edinburgh, UK

Running title: Genetic variants and colorectal cancer survival

Abbreviations list:

CRC, colorectal cancer; GWAS, genome-wide association study; *IQCM, IQ motif containing M gene;* SOCCS, Study of Colorectal Cancer in Scotland; HR, hazard ratio; CI, confidence interval; MAF, minor allele frequency; AJCC, American Joint Committee on Cancer.

***Corresponding authors:**

Malcolm G. Dunlop, Institute of Genetics and Molecular Medicine, The University of Edinburgh, Western General Hospital, Edinburgh EH4 2XU, UK, Tel +44 (0) 131 651 8631, Email address: malcolm.dunlop@ed.ac.uk

Evropi Theodoratou, Centre for Global Health Research, Usher Institute, University of Edinburgh, Teviot Place, Edinburgh, EH8 9AG, United Kingdom; Tel: (+44) 131-650-6194; E-mail: E.Theodoratou@ed.ac.uk

Conflict of Interest Statement: The authors declare no potential conflicts of interest.

Number of tables: 2

Number of references: 8

Abstract

Background: Germline genetic variants may influence pathways of tumor progression common to multiple cancer types. Here, we investigated the association between survival after colorectal cancer (CRC) diagnosis and 128 common genetic variants previously associated with prognosis in genome-wide association studies (GWAS) in different cancer types.

Methods: We studied survival outcomes in a large well-documented, prospective, population-based cohort (5,675 CRC patients) with up to 20 years follow-up.

Results: None of the 128 variants were significantly associated with overall or CRC-specific survival ($p<5x10^{-4}$, Bonferroni-corrected threshold). We observed suggestive evidence ($p<0.05$) for eight variants (rs17026425, rs17057166, rs6854845, rs1728400, rs17693104, rs202280, rs6797464, rs823920) in all CRC and two variants (rs17026425, rs6854845) in rectal cancer that were concordant with previous reports.

Conclusions: Given good statistical power (>0.80 for 75% of variants), this study indicates that most previously reported variants associated with cancer survival have limited influence on CRC prognosis.

Impact: Although small effects cannot be excluded, clinically meaningful germline influences on CRC patients as a group are unlikely.

Introduction

Colorectal cancer (CRC) is the second leading cause of cancer deaths worldwide(1). However, current knowledge on germline genetic influences over CRC prognosis is sparse. There is evidence that shared germline genetic basis exists across multiple cancer types in several key regulatory pathways of cancer pathogenesis(2) and progression(3). Previous genome-wide association studies (GWASs) have identified a number of genetic loci that might be associated with prognostic outcomes for various cancers. These genetic variants may also influence survival outcomes of CRC patients. Here, we report a large population-based study investigating the effects of published GWAS-identified variants associated with cancer prognosis on CRC survival.

Materials and Methods

We searched the NHGRI-EBI GWAS Catalogue (https://www.ebi.ac.uk/gwas/ accessed in December 2018) to retrieve GWAS identified variants ($p<5x10^{-5}$) associated with survival related traits for patients of any types of cancer. CRC patients with available information on age at diagnosis, sex, American Joint Committee on Cancer (AJCC) stage and GWAS data were included from the Study of Colorectal Cancer in Scotland (SOCCS). The MultiCentre Research Ethics committee for Scotland and other commitees approved the study and written informed consent was obtained from all participants. Additional details on the study cohort and quality control measures on genotyping have been reported previously (4, 5). CRC patients were prospectively followed up until death or censored on July 1st 2017. We evaluated overall survival (OS) and CRC-specific survival (CSS) as outcomes. A Cox proportional hazards model was adopted to estimate the effect of each variant (under an additive genetic model) on survival outcomes adjusting for age, sex and AJCC stage. We also

performed stratified analyses by sex, AJCC stage and tumor site. With the type I error at $\alpha < 5x10^{-4}$ (a Bonferroni corrected threshold), we estimated the study power for variants of various minor allele frequencies (MAF) and effect sizes using the method proposed by Owzar et al(6).

Results

A total of 5,675 CRC cases were included in this analysis and their basic characteristics are summarized in **Table 1**. One-hundred and twenty-eight genetic variants (linkage disequilibrium $r^2<0.2$) were identified from GWAS Catalogue (details presented in supplementary **Table S1**) and were included in the analysis. Power calculation indicated a power of at least 0.80 to detect a hazard ratio (HR) of 1.25 for 75% of the included variants (MAF>0.1). Power estimates with various parameters are presented in supplementary **Table S2**. In the overall analysis of all 5,675 CRC patients, none of the included variants were significantly associated with either OS or CSS (at $p<0.0005$). We observed eight variants (rs17026425, rs17057166, rs6854845, rs1728400, rs17693104, rs202280, rs6797464, rs823920) with $p<0.05$ in the same direction of effects with previous findings (**Table 2**); of them, three variants (rs17026425, rs17057166, rs6854845) were previously reported to be associated with rectal cancer survival. In stratified analysis, the variant rs17026425 was statistically significantly associated with OS for male CRC patients (HR=1.37, 95% CI=1.15-1.62, $p=3.3x10^{-4}$). Additionally, we observed two variants to be associated at $p<0.05$ (rs17026425, rs6854845) with OS in rectal cancer patients. No statistically significant associations were found in other stratified analyses (**Table S3-S5**).

Discussion

Here, we studied all common variants previously reported to be associated with prognosis in different cancer types. Overall, our results do not support any associations between these variants and survival outcomes for CRC patients. There are some suggestive signals that may merit further investigation in even larger datasets. For instance, we report a suggestive effect of rs17026425 in both overall and stratified analysis of rectal cancer patients, which concords with a previous GWAS(7). Of note, neither our study nor the previous GWAS detected association of this variant with colon cancer survival, indicating that this potential effect may be specific to rectal cancer. The variant is an intron variant of *IQ motif containing M (IQCM)* gene and is located in the binding region of JUN/JUND transcription factors, which manifest higher expression in CRC(8). The IQCM gene itself is highly expressed in testis only, making results restricted to males only in our study even more intriguing.

Presented here the study has sufficient power to detect 75% of previously reported survival variants, but failed to do so. Notably, 90% (19/21) of identified studies (**Table S1**) have sample size below 5,675 which is required to detect effect of genetic variants with MAF of 10% and HR of 1.25, thus suggesting potential false positive association as well as overestimation of real effects in original studies (winner's curse). Lack of pleiotropic and common effects across different cancers could also be behind the observed results, given the fact that variants reportedly associated with prognosis of other cancers except CRC showed mostly null effects in SOCCS. Our findings show poor reproducibility of results in the field and a pressing need for collaborative efforts, so as to aggregate larger CRC cohorts with genotype data to unravel the genetic architecture of CRC survival.

Acknowledgements

We acknowledge the excellent technical support from Stuart Reid and Marion Walker. We are grateful to Donna Markie and all those who continue to contribute to recruitment, data collection, and data curation for the Study of Colorectal Cancer in Scotland studies. We acknowledge that these studies would not be possible without the patients and surgeons who take part. We acknowledge the expert support on sample preparation from the Genetics Core of the Edinburgh Wellcome Trust Clinical Research Facility.

This work was supported by CRUK programme grant C348/A18927 (MGD). It was also supported by funding for the infrastructure and staffing of the Edinburgh CRUK Cancer Research Centre. ET is supported by a CRUK Career Development Fellowship (C31250/A22804). JPB is supported by an Edinburgh Clinical Academic Track-linked CRUK PhD Fellowship. PVS was supported by MRC Clinical Research Training Fellowship (MR/M004007/1), a research fellowship from the Harold Bridges Bequest and by the Melville Trust for the Care and Cure of Cancer. YH and XL were supported by the China Scholarship Council. The work received support from COST Action BM1206. This work was also funded by a grant to MGD as Project Leader with the MRC Human Genetics Unit Centre Grant (U127527202 and U127527198 from 1/4/18).

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2018;68:394-424.
2. Jiang X, Finucane HK, Schumacher FR, Schmit SL, Tyrer JP, Han Y, et al. Shared heritability and functional enrichment across six solid cancers. Nat Commun. 2019;10:431.
3. Hunter KW, Crawford NP. Germ line polymorphism in metastatic progression. Cancer Res. 2006;66:1251-4.
4. He Y, Timofeeva M, Farrington SM, Vaughan-Shaw P, Svinti V, Walker M, et al. Exploring causality in the association between circulating 25-hydroxyvitamin D and colorectal cancer risk: a large Mendelian randomisation study. BMC Med. 2018;16:142.
5. He Y, Theodoratou E, Li X, Din FV, Vaughan-Shaw P, Svinti V, et al. Effects of common genetic variants associated with colorectal cancer risk on survival outcomes after diagnosis: a large population-based cohort study. International journal of cancer. 2019 Jul 4. doi: 10.1002/ijc.32550. [Epub ahead of print].
6. Owzar K, Li Z, Cox N, Jung SH. Power and sample size calculations for SNP association studies with censored time-to-event outcomes. Genet Epidemiol. 2012;36:538-48.
7. Xu W, Xu J, Shestopaloff K, Dicks E, Green J, Parfrey P, et al. A genome wide association study on Newfoundland colorectal cancer patients' survival outcomes. Biomark Res. 2015;3:6.
8. Wang H, Birkenbach M, Hart J. Expression of Jun family members in human colorectal adenocarcinoma. Carcinogenesis. 2000;21:1313-7.

Table 1 Summarized characteristics of the SOCCS cohort

Basic characteristics	CRC cases (n=5,675)
Age at diagnosis (years)*	64.5(54.6-71.6)
Sex	
Male	3,235(57.0%)
Female	2,440(43.0%)
Site	
Colon	3,392(59.8%)
Rectum	2,201(38.8%)
Colon & rectum	16(0.3%)
Unknown	66(1.2%)
AJCC stage	
I	1,005(17.7%)
II	1,891(33.3%)
III	1,995(35.2%)
IV	784(13.8%)
No. of all-cause deaths	1,918(33.8%)
No. of CRC-related deaths	1,358(23.9%)

*Median and quartiles in parenthesis.

CRC, colorectal cancer; SOCCS, Study of Colorectal Cancer in Scotland; AJCC, American Joint Committee on Cancer.

Table 2 Summarized results of genetic variants that are associated with CRC survival (p<0.05) in SOCCS

Variant	locus	Gene	GWAS outcome that SNP was originally reported	Reported Effect(HR)	MA	MAF	Estimates in SOCCS		
							HR*(95%CI)	p-value	Power**
Overall survival									
rs17026425	4q31.23	IQCM	Rectal cancer(OS)	5.06	A	0.079	1.16(1.01-1.33)	0.039	1.00
rs17057166	5q33.3	LINC01847	Rectal cancer(DFS)	5.56	T	0.088	1.14(1.00-1.29)	0.042	1.00
rs6854845	4q13.3	Intergenic	Rectal cancer(DFS)	3.31	T	0.119	1.14(1.01-1.29)	0.040	1.00
rs1728400	16q24.1	Intergenic	Breast cancer(OS)	0.80	A	0.330	0.93(0.87-0.99)	0.026	1.00
rs17693104	10q23.1	SH2D4B	Serous epithelial ovarian cancer(OS)	1.65	T	0.348	1.08(1.01-1.15)	0.021	1.00
rs11138220	9q21.31	Intergenic	Rectal cancer(DFS)	2.76	G	0.131	0.88(0.79-0.98)	0.016	1.00
CRC-specific survival									
rs17693104	10q23.1	SH2D4B	Serous epithelial ovarian cancer(OS)	1.65	T	0.348	1.09(1.01-1.17)	0.031	1.00
rs202280	8q21.13	intergenic	Serous epithelial ovarian cancer(OS)	2.00	G	0.038	1.14(1.02-1.26)	0.018	1.00
rs6797464	3q26.2	MECOM	Osteosarcoma(OS)	1.80	A	0.119	1.18(1.02-1.37)	0.030	1.00
rs823920	9q31.1	Intergenic	Pancreatic cancer(OS)	1.43	G	0.123	1.11(1.00-1.23)	0.042	1.00
rs11138220	9q21.31	Intergenic	Rectal cancer(DFS)	2.76	G	0.131	0.85(0.75-0.97)	0.016	1.00

*Hazard ratios are estimated based on minor alleles.

** Statistical power is estimated using originally reported effect sizes with type I error (α) at 0.0005.

CRC, colorectal cancer; SOCCS, Study of Colorectal Cancer in Scotland; GWAS, genome-wide association studies; OS, overall survival; DFS, disease-free survival; MA, minor allele. MAF, minor allele frequency.